

DOCUMENT RESUME

ED 241 548

TM 840 013

AUTHOR Carstens, Paul W.; McKeag, Robert A.
TITLE The Effect of a Change in Item Sequence Order on Performance in a Test, Re-Test Experiment.
PUB DATE [82]
NOTE 9p.
PUB TYPE Reports - Research/Technical (143)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Higher Education; Objective Tests; *Performance Factors; Scores; *Test Format; *Testing Problems; Test Results
IDENTIFIERS *Item Position (Tests); Standard Deviation

ABSTRACT

This study utilized a test, re-test procedure to investigate what effects a change in the order, or sequence, of test items would have on student performance. College juniors (n=102) were given a 50-item multiple-choice and matching item test on the general subject of educational measurement. The items had no particular sequence, but had simply been taken at random from an item bank on the subject. Two weeks later the same students were retested; alternate rows of students received the experimental version, which differed from the original only in item sequence. It was expected that a learning effect would be present, resulting in higher scores and smaller standard deviations on the re-test. Results only partially aligned with expected results. While mean performance was higher on the re-test results, standard deviation did not follow expected results. This study concludes that the sequencing of items in a test is a factor which may be more complex than previously thought and needs careful study and consideration by teachers and test makers. (Author/BW)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

ED241548

The effect of a change in item sequence order on performance in
a test, re-test experiment

Paul W. Carstens
Robert A. McKeag
University of Wisconsin-Oshkosh
Oshkosh, WI 54901

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

✓ This document has been reproduced as
received from the person or organization
originating it.

Minor changes have been made to improve
reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

P.W. Carstens

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC).

ABSTRACT

This study utilized a test, re-test procedure to investigate what effects a change in the order, or sequence, of test items would have on student performance.

Results only partially aligned with expected results. While mean performance was higher on the re-test results, standard deviation did not follow expected results.

This study concludes that the sequencing of items in a test is a factor which may be more complex than previously thought and needs careful study and consideration by teachers and test makers.

Introduction

The notion of sequence as a factor in teaching and learning and testing has long been recognized as sometimes critical. In some curricular areas, as, for example, in one approach to foreign language instruction where dialogs are memorized, it has been argued that the dialogs should be learned and tested and that the original sequence is an integral part of the "meaning" of the dialog. In the area of reading instruction also, students are often asked to note that concepts are represented by chapters in the book. If such concepts are assumed to have necessary sequences and the concepts in the chapter would lose their "integrity" if the elements of that concept were rearranged, then it would follow that attention must be given to maintaining this sequence for both testing and learning situations. It has also been suggested that students tend to learn that for which they are tested -- that, "it is the evaluation structures that determine the functional objectives of instruction" (Gronlund, 1976). If this is the case and the sequence, or arrangement of items in a message, is critical, then it may be argued that this sequence should be maintained for any testing done on the "message" or concept involved.

The test, re-test procedure has been used previously to investigate test reliability (Chase, 1978) or in studies of test construction (Cangelosi, 1982). Here the procedure was used to determine possible order or sequence effects on student test performance.

The Study

Subjects for this study were 102 junior level uni-

versity students who were taking reading courses in a college of education. All had previously taken a standard "tests and measures" course in a teacher education program.

The study was conducted in three different test administrations over the course of two semesters. For each administration, students were asked to participate in a test, re-test experiment and then were given a 50 item investigator-designed multiple-choice and matching item test on the general subject of educational measurement. No student declined to participate though that was a possible option. The items on the test had no particular sequence, but had simply been taken at random from a data bank of test items dealing with the first unit in the tests and measures course.

Two weeks later the same students were retested as in the first administration except that alternate rows of students received experimental versions of the test, i.e. the experimental version differed from the original in that the sequence of items was different. That was the only difference between the original test and the experimental version. Directions to students and individual items were the same for each administration.

Results

The results show that for two out of three re-tests retaining the original sequence of items (TA), means are higher than for the first test administration while Sd's were smaller on the re-tests for two out of three cases. Where the item sequence differed randomly on the re-test, i.e. the experimental version (TX), however, mean performance was always higher

than the means on the original test and the Sd's were also consistently higher for the re-test using a different item sequence.

When the re-tests TA and TX were compared, the mean scores were higher in all three administrations for TX (item sequence changed) than for TA (original item sequence retained). The Sd's for TX were also higher than for TA in all three administrations.

Results are summarized below.

<u>TEST</u>		<u>RE-TEST(TA, original sequence)</u>	<u>RE-TEST(TX, changed seq)</u>
x	30.84	32.33	33.33
Sd	6.76	6.15	8.40
N	38	19	19
x	31.00	31.85	32.31
Sd	6.56	4.79	7.54
N	30	15	15
x	26.33	25.42	28.67
Sd	6.00	6.63	8.03
N	34	17	17

When a t-test on mean difference between TA and TX results was run, differences were found to be non-significant largely due, it was felt, to the small N's which result when the test group was divided on the re-test. If one is willing to pool three test administrations results mean differences between TA and TX are significant at the .05 level.

Discussion

It would be expected that a learning effect would be present in a test, re-test situation since the individual

test items were the same. It would also be expected that standard deviations would be smaller on a re-test because of this learning effect. In general this was true for the results of this experiment if we look only at the test, re-test (TA, original item sequence retained) data.

The general expectation of higher scores and smaller standard deviations was not borne out, however, in this experiment when we look at data from the test, re-test (TX, item sequence changed). Here performance was higher than the original test and the TA re-test but standard deviations were larger than for all other test administrations.

It is not easy to account for the discrepancy between expected performance and actual performance noted in the test, re-test (TX) data.

Further, it should be observed that there was no inherent sequence of items in the original test or any particular hierarchy of concepts or "meaning." The sequence factor was simply added into the TX re-test situation and simply varied the sequence of items randomly from the rather casual sequence of items in the original test. It would seem that this should make any sequencing effect less noticeable. The fact that possible sequence effects were in fact more noticeable in the TX results could be interpreted to mean that such effects are even more powerful than suspected.

It is possible to postulate a "loss of interest" or a "loss of focus" factor, especially for TA results, to account for some results. An examination of the last ten items in the test gave some indications that students may not have attended to the end of the test as well for TA as they did for TX.

Possibly students doing the TA re-test recognized the test as one they had taken two weeks earlier and tended to get careless and this did not happen as readily for students doing the TX re-test because of the changed sequence of items.

While we feel that there are some possible implications from this study we wish to be modest in any discussion in view of the small N's involved in the re-test data. It would probably be useful to replicate this study with a larger group to see if these results can be repeated. It certainly seems to be clear that the sequencing of items in a test can have some effect on student performance even when there is no apparent hierarchy or sequence to the "meaning" of the subject matter. Whether the re-test results in this experiment are due to an "interest" factor or the change in item sequencing factor or some other, unrecognized variable, it does appear to us that this study has demonstrated that the sequence factor in learning and testing is worthy of further investigation. It also appears to be a more subtle, complex factor than many teachers have thought. These possibilities need to be explored and considered carefully by teachers and test makers.

References

Cangelosi, James. 1982. Measurement and evaluation. Dubuque, IA: Wm. C. Brown.

Chase, Clinton. 1978. Measurement for educational evaluation. Menlo Park, CA: Addison Wesley.

Gronlund, Norman. 1966. Measurement and evaluation in teaching. New York: Macmillan.